Case-Finding Instruments for Depression in Primary Care Settings

Cynthia D. Mulrow, MD, MSc; John W. Williams Jr., MD, MHS; Meghan B. Gerety, MD; Gilbert Ramirez, DrPH; Oscar M. Montiel, MD; and Caroline Kerber, MD

Objective: To evaluate the usefulness of case-finding instruments for identifying patients with major depression in primary care settings.

Data Sources: A MEDLINE search of the English-language medical literature; bibliographies of selected papers; and experts.

Study Selection: Studies that were done in primary care settings with unselected patients and that compared case-finding instruments with accepted diagnostic criterion standards for major depression were selected.

Data Synthesis: 9 case-finding instruments were assessed in 18 studies. More than 15,000 patients received screening with a case-finding instrument; approximately 5300 of these received criterion standard assessment. Case-finding instruments ranged in length from 2 to 28 questions. Average administration times ranged from less than 2 minutes to 6 minutes. Sensitivities and specificities for detecting major depression ranged from 67% to 99% and from 40% to 95%, respectively. No significant differences between instruments were found. Overall sensitivity was 84% (95% CI, 79% to 89%); overall specificity was 72% (CI, 67% to 77%). If a case-finding instrument were administered to 100 primary care patients with a 5% prevalence of major depression, the clinician could expect that 31 patients would screen positive, that 4 of the 31 would have major depression, and that 1 patient with major depression would not be identified.

Conclusions: Several instruments with reasonable operating characteristics are available to help primary care clinicians identify patients with major depression. Because the operating characteristics of these instruments are similar, selection of a particular instrument should depend on issues such as feasibility, administration and scoring times, and the instruments’ ability to serve additional purposes, such as monitoring severity or response to therapy.

Depressive disorders are common, persistent, and recurring afflictions among primary care patients. They cause substantial suffering for patients and their families and are associated with a loss of personal productivity and a markedly increased risk for suicide. Further, the presence of depression puts persons with comorbid conditions, such as recent myocardial infarction, at increased risk for illness and death. Persons with depression spend more time with their physicians during office visits and use more health care than persons without depression (1). The annual health care cost associated with depression in the United States is estimated to be $43.7 billion (2). Under-scoring the importance of identifying patients with depression is that the effectiveness of therapy, including antidepressants, psychotherapy, and counseling, has clearly been established (3).

Despite these issues, primary care providers fail to diagnose and treat as many as 35% to 50% of patients with depressive disorders (4, 5). Obstacles to the appropriate recognition of depression include inadequate provider knowledge of diagnostic criteria; competing comorbid conditions and priorities among primary care patients; time limitations in busy office settings; concern about the implications of labeling; poor reimbursement mechanisms; and uncertainty about the value, accuracy, and efficiency of screening mechanisms for identifying patients with depression. We address the last of these obstacles and assess the feasibility and operating characteristics of several case-finding instruments that have been used to detect depressive disorders in primary care settings. Our ultimate goal is to familiarize providers with the advantages and disadvantages of these instruments so that they can make informed decisions about incorporating them into practice.

Methods

Data Acquisition

We did a MEDLINE search of the English-language medical literature published from 1966 through February 1994. Search terms included “depressive disorder or depression,” “diagnosis,” and the specific names of each of 11 case-finding instruments cited in previous relevant reviews or bibliographies (6-9). Other sources were references identified from pertinent articles and national experts in the field of depression. Experts included authors of papers that were selected for review and two members of the Agency for Health Care Policy and Research Guideline Panel on Depression.

Of 906 articles identified through MEDLINE, 210 were deemed potentially relevant. These were reviewed to identify studies that met the following selection criteria.

Study samples had to have been composed of primary care patients attending clinic. Patients were excluded if they had been selected because they had specific conditions (such as chronic pain or cancer) or because they had specific demographic char-
Table 1. Characteristics of Case-Finding Instruments That Have Been Used to Detect Depression in Primary Care Settings*  

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Scope</th>
<th>Items, n</th>
<th>Response Format</th>
<th>Time Frame of Questions</th>
<th>Score Range</th>
<th>Usual Cut-point‡</th>
<th>Literacy Level§</th>
<th>Administration Time, min</th>
<th>Monitor Severity or Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI</td>
<td>Depression-specific</td>
<td>21</td>
<td>4 statements of symptom severity per item</td>
<td>Today</td>
<td>0–63</td>
<td>10 mild</td>
<td>20 moderate</td>
<td>30 severe</td>
<td>2–5</td>
</tr>
<tr>
<td>CES-D</td>
<td>Depression-specific</td>
<td>20</td>
<td>4 frequency ratings: “less than one day” to “most or all (5–7) days”</td>
<td>Past week</td>
<td>0–60</td>
<td>16</td>
<td></td>
<td></td>
<td>2–5</td>
</tr>
<tr>
<td>GHQ</td>
<td>Psychiatric illness (several versions)</td>
<td>28</td>
<td>4 frequency ratings: “not at all” to “much more than usual”</td>
<td>Past few weeks</td>
<td>0–28</td>
<td>4</td>
<td>Easy</td>
<td>5–10</td>
<td>Yes</td>
</tr>
<tr>
<td>HSCL</td>
<td>Multiple versions and multiple components with depression category</td>
<td>25</td>
<td>4 frequency ratings: “not at all” to “much more than usual”</td>
<td>Past week</td>
<td>25–100</td>
<td>43</td>
<td>Average</td>
<td>2–5</td>
<td>Yes</td>
</tr>
<tr>
<td>MOS-D</td>
<td>Depression-specific</td>
<td>8</td>
<td>Frequency ratings; same format as CES-D</td>
<td>Past week</td>
<td>0–1 (logistic regression)</td>
<td>0.06</td>
<td>Average</td>
<td>&lt;2</td>
<td>Yes</td>
</tr>
<tr>
<td>ID</td>
<td>Depression-specific</td>
<td>15</td>
<td>2 items, yes or no; 3 statements of normal, overt, or covert symptomatology</td>
<td>Recently</td>
<td></td>
<td></td>
<td>Easy</td>
<td>2–5</td>
<td>Unknown</td>
</tr>
<tr>
<td>PRIME-MD</td>
<td>Multiple components with depression category</td>
<td>2</td>
<td>Yes or no</td>
<td>Past month</td>
<td>0–2</td>
<td>1</td>
<td>Average</td>
<td>&lt;2</td>
<td>Unknown</td>
</tr>
<tr>
<td>SDDS-PC</td>
<td>Multiple components with depression category</td>
<td>5</td>
<td>Yes or no</td>
<td>Past month</td>
<td>0–4</td>
<td>2</td>
<td>Easy</td>
<td>&lt;2</td>
<td>Unknown</td>
</tr>
<tr>
<td>SDS</td>
<td>Depression-specific</td>
<td>20</td>
<td>4 frequency ratings: “little of the time” to “most of the time”</td>
<td>Recently</td>
<td>25–100 (sum/80 X 100)</td>
<td>50 mild</td>
<td>60 moderate</td>
<td>70 severe</td>
<td>2–5</td>
</tr>
</tbody>
</table>

* BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HSCL = Hopkins Symptoms Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System—Primary Care; SDS = Zung Self-Assessment Depression Scale.

Item numbers for the PRIME-MD and SDDS-PC refer to depression questions only; item numbers for the HSCL refer to depression plus anxiety questions.

‡ Cut-point is the number at or above which the test is considered positive.
§ Easy = 3rd- to 5th-grade reading level; average = 6th- to 9th-grade reading level according to Fog Formula (38).

Characteristics (for example, they were immigrants in a particular ethnic group). Both a case-finding instrument and a diagnostic criterion standard had to have been administered. The criterion standard had to have had formal standardized diagnostic criteria for depression. Accepted criterion standards were the Diagnostic Schedule Manual-3 criteria (DSM-III or DSM-III-R) and the Research Diagnostic Criteria, or a close approximation of these. Standard interview procedures, such as the Diagnostic Interview Schedule or the Structured Clinical Interview for DSM-III, had to have been used to arrive at the diagnosis. Chart or physician diagnoses of depression that were made without specified formal interview procedures and diagnostic criteria were excluded.

Nineteen studies involving nine case-finding instruments met the selection criteria: Fourteen were found during the MEDLINE search; 1 came from a relevant bibliography; and 4 were unpublished at the time of the search and came from experts (10–28). Of the remaining articles screened, 92% were excluded because they did not involve primary care patients, 6% were excluded because they had used an inadequate criterion standard, and 2% were excluded because they involved selected populations (29, 30) or because they had tested modified and unvalidated versions of case-finding instruments (31).

Data Extraction

Articles were abstracted by two independent reviewers. Determination of study quality was made on the basis of 1) whether the case-finding instrument was administered and interpreted independently of the criterion standard and 2) whether the proportion of persons receiving the criterion standard assessment was less than or more than 50% of those approached for criterion standard assessment. Quality assessment addressed methodologic issues relevant to the evaluation of diagnostic tests (such as independent assessment and selection bias) and did not necessarily reflect the ability of studies to address their original aims. There were no disagreements about quality assessments.

Data Synthesis

Established cut-points for case-finding instruments (Table 1) were used. Two-by-two tables were constructed that categorized numbers of screened-positive and screened-negative persons who did and did not meet criterion standard diagnosis for major depression and major depression or dysthymia. Kraemer’s method (32) was used to adjust for verification bias for studies that used two-stage assessment techniques; whereby the criterion standard was administered only to a random sample of persons who screened negative on case-finding instruments (33). The authors of all but one study provided us with additional data and analyses when two-by-two tables could not be derived from abstraction of the published article. This one study (28) was dropped from further review because its authors could not be contacted and tables could not be derived from published information.
A scattergram (Figure 1) plotting true-positive against false-positive rates was constructed to visually evaluate variability among studies (34). To provide a visual reference for the consistency of study results, we modeled a summary receiver-operating curve based on the logit transformations of the true-positive and false-positive rates.

Average sensitivities and specificities, weighted by study size and corrected for two-stage assessment techniques when indicated, were computed both by case-finding instrument and by overall instruments (35). The point estimates and 95% CIs were calculated using a linear random-effects model (36, 37). Approximate 95% CIs were estimated using quadratic root formulae because most of the point estimates were near unity (37). Differences in weighted average sensitivities and specificities between case-finding instruments were evaluated using the z statistic with the Scheffe multiple-comparison adjustment (37). Stratified analyses were done to evaluate whether estimated sensitivities and specificities varied between high-quality studies and those with major selection bias or lack of independent assessment. Regression analysis was used to determine associations between reported study prevalences and sensitivity estimates (37).

Results

Descriptions of Case-Finding Instruments

Characteristics of the nine case-finding instruments that have been evaluated in primary care settings are presented in Table 1. All of the questionnaires are written either at the easy (3rd to 5th grade) or average (6th to 9th grade) reading level (38). Almost all can be self-administered in less than 5 minutes. Except for the General Health Questionnaire, all include specific questions aimed at assessing depressed mood or whether a patient feels sad or blue. All include questions assessing anhedonia. Most are available in languages other than English, such as Spanish.

The Beck Depression Inventory, the Center for Epidemiologic Studies Depression Screen, and the Zung Self-Assessment Depression Scale are three commonly used, traditional instruments that were developed specifically to identify depression. They include similar numbers of questions and use response formats that rely either on ranking symptom severity or on classifying frequency of symptoms. The time frames of questions are “today” for the Beck Depression Inventory, “over the past week” for the Center for Epidemiologic Studies Depression Screen, and “recently” for the Zung Self-Assessment Depression Scale. These three instruments have been used in numerous settings (including the community, the clinic, and the hospital) not only to identify depression but also to rate severity of depression and to monitor response to therapy.

| Table 2. Studies Evaluating Case-Finding Instruments for Depression in Primary Care Settings* |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Year | Study (Reference) | Instrument | Sample Size† | Population‡ | Criterion Standard | Quality Score§ |
| 1990 | Perez-Stable et al. (10) | BDI | 265 | Mixed | DSM-III | 4 |
| 1990 | Zich et al. (11) | CES-D | 214 | Mixed | DSM-III | 4 |
| 1993 | Kirmayer et al. (12) | CES-D | 685 | Academic | DSM-III | 1 |
| 1983 | Schulberg et al. (13) | CES-D | 294 | Community | DSM-III | 4 |
| 1994 | Hendrie et al. (14) | CES-D | 125 | Academic | DSM-III-R | 2 |
| 1990 | Hough et al. (15) | CES-D | 525 | (age > 59 years) | HMO | DSM-III | 2 |
| 1994 | Fechner-Bates et al. (16) | CES-D | 425 | Community | DSM-III-R | 2 |
| 1979 | Hooper et al. (17) | CES-D | 247 | Academic | RDC | 2 |
| 1987 | Von Korff et al. (18) | CES-D | 809 | Academic | DSM-III | 1 |
| 1987 | Goldberg and Blackwell (19) | CES-D | 283 | Community | DSM-III | 1 |
| 1977 | Finlay-Jones and Murphy (20) | CES-D | 100 | Community | DSM-III-R-like algorithm | 2 |
| 1982 | Okimoto et al. (21) | ID | 55 | Veterans Affairs | DSM-III | 3 |
| 1994 | Williams et al. (22) | MOS-D | 99 | Mixed | DSM-III-R | 2 |
| 1988 | Burnam et al. (23) | MOS-D | 501 | HMO | DSM-III | 4 |
| 1994 | Broadhead et al. (24)† | MOS-D | 388 | Community | DSM-III-R | 3 |
| 1994 | Broadhead et al. (24)‡ | SDDS-PC | 257 | Mixed | DSM-III-R | 3 |
| 1994 | Spitzer et al. (25) | PRIME-MD | 431 | Mixed | DSM-III-R | 1 |
| 1990 | Magruder-Habib et al. (26) | SDDS-PC | 206 | Veterans Affairs | DSM-III | 4 |
| 1979 | Raft et al. (27) | SDDS-PC | 69 | Academic medical | DSM-III-like | 4 |

* BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HMO = health maintenance organization; HSCL = Hopkins Symptom Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System-Primary Care; SDS = Zung Self-Assessment Depression Scale.
† The sample size refers to the actual number who received the criterion standard.
‡ Mixed-community and university-affiliated clinics; academic-university-affiliated clinics; community-private practice clinics.
§ Lower scores indicate higher quality.
† This study used the same sample as that used in the study by Hough and coworkers (15).
The General Health Questionnaire and the Hopkins Symptom Checklist are questionnaires that screen for general psychiatric illness; the Hopkins Checklist has a specific category for depression. Both of these instruments have several versions with different numbers of questions. The Medical Outcomes Study Depression Screen is a depression-specific screening instrument that was developed by combining two questions from the Diagnostic Interview Schedule (39) with six questions from the Center for Epidemiologic Studies Depression Screen. A logistic regression scoring method is used; this requires a calculator.

The Primary Care Evaluation of Mental Disorders (PRIME-MD) and the Symptom Driven Diagnostic System-Primary Care instruments are recently developed, multidimensional questionnaires. Each has screening questions arranged in several categories (for example, mood or depression, anxiety, alcohol abuse, and somatization) that are used to trigger more extensive diagnostic interviewing sections for specific DSM-III-R diagnoses. The depression components of these two instruments include the fewest questions of all case-finding instruments that have been studied in primary care settings.

Descriptions of Studies and Findings

The 18 studies that we reviewed are outlined in Table 1. The cut-points for instruments were used for these calculations (Table 1). The cut-point for mild depression was used for the two scales with three listed cut-points (the Beck Depression Inventory [10] and the Zung Self-Assessment Depression Scale [50]; the study by Raft and coworkers (27) only had information on a cut-point of 60 for the Zung Scale. The scattergram suggests that one study varies systematically from the others (27); this was the small study by Raft and coworkers, which used the higher cut-point for the Zung Scale. This study had an unusually low sensitivity (31%; CI, 16% to 51%) for detecting major depression.

Figures 2 and 3 show sensitivities and specificities, respectively, for detecting major depression using study and case-finding instruments. The average weighted sensitivity of all studies, except for that by Raft and coworkers (27), was 84% (CI, 79% to 89%); the average weighted specificity was 72% (CI, 67% to 77%). There were no statistically significant differences for either sensitivities or specificities among the different case-finding instruments ($P > 0.05$).

Only seven studies had data on the ability of instruments to identify persons with either dysthymia or major depression (10, 13, 15, 18, 23–25). Operating characteristics for detecting this combined category were not statistically significantly different from those for detecting major depression alone, although slightly lower sensitivity and higher specificity were seen. Average weighted sensitivity was 78% (CI, 73% to 83%); average weighted specificity was 77% (CI, 71% to 82%). Stratified analyses were done to evaluate whether sens-
sensitivities and specificities derived from studies with selection bias and from studies that did not assess criterion standard diagnosis independently of case-finding score differed from sensitivities and specificities derived from higher-quality studies. Sensitivities ranged from 79% to 83%; specificities ranged from 66% to 74%. Analyses stratified according to whether the DSM-III (11 studies) or DSM-III-R (6 studies) criterion standard was used showed sensitivities of 84% and 83%, respectively, and specificities of 69% and 76%, respectively. Studies that examined current depression (depression within the previous 2 to 4 weeks) had a sensitivity of 84% and a specificity of 75%, and those that either examined depression over the previous year or were unclear about their diagnostic time frame had a sensitivity of 86% and a specificity of 62%. None of the stratified analyses discussed above showed statistically significant differences between groups. Finally, regression analyses showed no significant associations between reported study prevalences of major depression and sensitivity estimates.

### Expected Yield of Case-Finding Instrument Strategy

Figure 4 shows what a clinician could expect if he or she were to use a case-finding instrument in a clinic setting. This scenario assumes that the clinician sees 100 adult patients (a reasonably booming outpatient practice for a week) and that the prevalence of major depression in the clinic is 5% (a prevalence commonly found in primary care settings). If the sensitivity of the instrument were 80% and the specificity were 72% (Figures 2 and 3), 31 patients would score at or above the usual cut-point for the case-finding instrument and would be considered possibly depressed. Of these, 4 would meet the DSM-III-R criteria for major depression. Among the 69 patients who screened negative, 1 would actually have major depression. If the prevalence of major depression in the clinic setting was as high as 9%, 7 of the patients who screened positive would be correctly identified and 2 patients with major depression would not be diagnosed.

The above-mentioned sensitivity and specificity values convert to a likelihood ratio of 2.86 (sensitivity / specificity). This likelihood ratio can be used to determine the post-test probability of a patient having major depression if he or she scores at or above the cut-point of the case-finding instrument. If a clinician practices in a setting where the prevalence or pretest probability of depression is 5%, a positive screening would increase the probability that a patient is suffering from major depression to 13%. If the prevalence in the clinician’s setting is 9%, then a positive screening would increase the probability of major depression to 20%.

### Discussion

Depression is one of the most common psychiatric diagnoses in primary care populations; major depressive disorders can be diagnosed in 6% to 9% of such patients (40). Depression is associated with significant burden and costs, and effective treatment is clearly available. Given that 50% to 60% of persons seeking help for depression are treated exclusively in the primary care setting, accurate detection in this setting is important (41). One strategy that may improve detection is the use of psychiatric screening or case-finding instruments in the physician’s practice.
Figure 3. Specificities of case-finding instruments (CFI) for identifying major depression. Point estimates with 95% CIs are presented for single studies (*), for summary data on case-finding instruments (•), and overall (○). BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HSCL = Hopkins Symptoms Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System-Primary Care; and SDS = Zung Self-Assessment Depression Scale.

office as part of the routine health evaluation. In contrast to screening applied to unselected populations, case-finding occurs when clinicians search for disease by screening their own patients. Several things should be considered before this strategy is chosen.

First, are there case-finding instruments that are feasible to use in the primary care setting? The answer is yes. Several case-finding instruments are relatively short and can be easily self-administered by patients with at least grade-school to middle-school reading ability. For patients without such reading skills, instruments can be administered by trained office staff. Several of the instruments are available in different languages, including Spanish. All are easily scored, except for the Medical Outcomes Study Depression Screen, which requires a calculator.

Second, do the operating characteristics of the case-finding instruments enable the clinician to decrease the number of persons who need more intensive diagnostic interviewing without missing persons who might potentially benefit from a full diagnostic interview? The answer is probably yes. We have shown that in a group of 100 patients in a clinic with an average prevalence of 5% for major depression, the clinician would have to do a diagnostic interview in only 31 patients to identify the 4 with major depression, and only 1 person with major depression would be missed. It is unlikely that many busy clinicians could do or would want to do diagnostic interviews in all 100 patients so that they would not miss 1 case of depression. Further, when using a case-finding instrument, the clinician can expect to miss about 20% of patients with depression; current practice often misses 50% of such patients (4, 5).

Third, is a case-finding strategy effective in identifying patients with major depression who will receive treatment and who will have improved clinical outcomes? The answer is that we don’t know. Although we did not directly address this question, at least 12 studies have assessed the effect of feedback of psychiatric case-finding scale scores on physician practice patterns (42-53). Some of these have shown improved recognition or treatment of depression with such feedback, or both, but none have shown improved clinical outcomes. Most of the studies had significant design limitations. Only 1 incorporated a multifaceted approach to help ensure that clinicians had adequate knowledge with which to make accurate depression diagnoses and treatment recommendations (42). Regardless, the lack of data from randomized trials has led the American Academy of Family Physicians, the Canadian Task Force on the Periodic Health Examination, and the U.S. Preventive Services Task Force to not recommend the routine use of case-finding instruments (54-56). They recommend only that clinicians remain alert for symptoms of depression, particularly in high-risk patients. The recent Agency for Health Care Policy and Research Guideline Panel on Depression noted that case-finding instruments were unlikely to add much if clinicians were already highly attuned to and inquired regularly about depressive symptoms (3). The Panel also noted that the efficiency of case-finding instruments could be increased by targeting high-risk patients.

Unfortunately, recent data suggest that primary care clinicians still miss as many as 35% to 65% of patients with major depression (4, 5, 10, 57). Clinician recognition may be unaffected by a patient’s sex or level of education,
may be slightly lower among younger patients, and may be higher among patients with the most severe and disabling disease (4, 57, 58). Fewer than 20% of persons with major depression in the primary care setting have severe disease, and a sparse amount of available data suggests that 25% of these persons are still being missed (57). Although case-finding instruments detect a higher proportion of persons with major depression (80%) than primary care clinicians do (35% to 65%), it is unclear whether they miss the same spectrum of disease most commonly missed by clinicians. Because sensitivity tends to increase with increasing severity of disease, it is likely that case-finding instruments will most often miss persons with mild disease and that they will not miss as many as 25% of persons with severe disease (59).

Targeting high-risk patients with a higher underlying prevalence of depression for case-finding is an intuitively appealing strategy because it offers greater efficiency by improving the positive predictive value of the case-finding instrument. However, gains in efficiency are related to the magnitude of the risk factor and the prevalence of that risk factor in the patient population (60, 61). For example, the strongest demographic risk factor for depression is female sex (relative risk, 2) (62, 63). In primary care practices, women patients outnumber men by a ratio of 60:40 (64). If only women were screened with a case-finding instrument that was 80% sensitive and 72% specific, the positive predictive value would increase only marginally, from 13% to 16%. This minimal gain would be associated with a 20% decrease in sensitivity for the case-finding strategy and a potential for missing 40% of depressed patients. A more potent (relative risk, 3) and less common (10%) risk factor is known previous depressive illness (65). If the “highly attuned” clinician limited case-finding to patients with this risk factor, he or she would increase the positive predictive value of the case-finding strategy to 29%, but overall sensitivity would be greatly reduced. Seventy-five percent of patients with major depression and no past known depressive illness would not be screened.

Incorporating case-finding instruments for depression into practice has its pitfalls. Routine screening for depressive disorders may lead clinicians to focus only on depression and ignore other common psychiatric disorders. Patients with anxiety disorders and alcohol abuse may have positive scores on depression instruments. Unless these disorders are ruled out, patients may be mistakenly labeled as depressed or may not be recognized as having depression comorbid with another psychiatric condition. Some case-finding instruments, such as the General Health Questionnaire, are designed to detect any psychiatric disorder and may be less likely to bias the clinician toward focusing only on depression. Two other recently developed instruments, the Primary Care Evaluation of Mental Disorders and the Symptom Driven Diagnostic System, are multidimensional questionnaires that assess several specific psychiatric disorders (24, 25). For example, the Primary Care Evaluation of Mental Disorders includes questions related to depression, anxiety, somatization disorder, and alcohol abuse. Positive responses to any of the four sections trigger specific structured interview questions that are used to make definitive diagnoses. Such multidimensional instruments may minimize undue emphasis on depression and identify comorbid psychiatric disorders.

Several limitations of our review merit mention. As with any review, ours represents a retrospective compilation of data, some of which were originally collected for reasons other than evaluating the accuracy of a case-finding instrument. In the studies, many fewer patients received criterion standards than were screened with case-finding instruments. Whether selection biases were present and what their potential effects may have been could not be evaluated. It is possible that persons with severe depression more often refused or were less likely to volunteer for criterion standard assessment. This could have resulted in a relatively healthy sample without adequate representation of persons with severe depression; this, in turn, would lead to an underestimation of sensitivity and an overestimation of specificity. On the other hand, if persons with severe depression were more likely to volunteer for criterion standard assessment, sensitivity would be overestimated and specificity underestimated. Although a formal criterion standard for the diagnosis of depression was used in all studies, these standards sometimes varied in required duration and in numbers and types of symptoms counted toward the diagnosis. However, stratified analyses did not show differences in the performance characteristics of different criterion standards. Lastly, the ability of instruments to identify related and clinically relevant diagnoses other than major depression or dysthymia, such as minor depression or anxiety, was not evaluated.

Despite these limitations, salient lessons that can be used to guide current decisions and frame future research are apparent. Nine different instruments have been used feasibly in primary care settings and have shown reasonable and consistent accuracy. Moreover, there is no need to develop new instruments aimed at identifying primary care patients with major depression or dysthymia. However, an important remaining question is whether any of
The authors thank the many investigators who provided depressive disorders, because most diagnostic schemes for these disorders require the presence of either anhedonia in his patients and asks additional questions only in patients with at least one of these two symptoms? (Such a strategy would theoretically identify most patients with depressive disorders, because most diagnostic schemes for these disorders require the presence of either depressed mood or anhedonia.) And a final question that continues to deserve study is, even if case-finding instruments are more acceptable, accurate, and efficient than routine physician questions during office visits, which strategy leads to better treatment and outcomes?

An informed, rational clinician can incorporate several feasible, reasonably accurate case-finding instruments for major depression into practice. Whether the use of such instruments ultimately will lead to improved outcomes is unknown. Choices about whether to use an instrument should take into account estimates of the resources required (for example, questionnaire forms, staff time for scoring, and for interviewing patients unable to complete self-administered forms) compared with potential resources saved (for example, diagnostic interviewing time, potentially improved outcomes, and decreased health care utilization if improved diagnosis is linked with treatment). If the clinician opts to use a case-finding instrument, several are available and have similar operating characteristics. Choice of a particular instrument should take into account ease of use and the time required for administration as well as whether the clinician wants to use the instrument for other purposes, such as screening for other psychiatric disorders or monitoring treatment response. Finally, regardless of whether primary care clinicians routinely incorporate case-finding instruments into their practices, all of them should know how to diagnose major depression and should aim to link accurate diagnosis with effective treatment strategies.

Acknowledgments: The authors thank the many investigators who provided either raw data or reanalysis of data or who made unpublished information available.

Grant Support: In part by a Robert Wood Johnson Generalist Physician Faculty Award and the Mexican American Medical Treatment Effectiveness Research Center funded by the Agency for Health Care Policy and Research.

Requests for Reprints: Cynthia D. Mulrow, MD, MSc, Audie L. Murphy Memorial Veterans Hospital (11C6), 7400 Merton Minter Boulevard, San Antonio, TX 78284.

Current Author Addresses: Drs. Mulrow, Williams, Gerety, Ramirez, Montiel, and Kerber: Audie L. Murphy Memorial Veterans Hospital (11C6), 7400 Merton Minter Boulevard, San Antonio, TX 78284.

References
34. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et